

# Conditional Spectral Analysis of Replicated Multiple Time Series with Application to Nocturnal Physiology

Robert T. Krafty, Ori Rosen, David S. Stoffer,  
Daniel J. Buysse, and Martica H. Hall \*

December 29, 2016

---

\*R. T. Krafty is Associate Professor, Department of Biostatistics, D. S. Stoffer is Professor, Department of Statistics, and D. J. Buysse and M. H. Hall are Professors, Department of Psychiatry, University of Pittsburgh (rkrafty@pitt.edu, stoffer@pitt.edu, buyssedj@upmc.edu, hallmh@upmc.edu). O. Rosen is Professor, Department of Mathematical Sciences, University of Texas at El Paso (ori@math.utep.edu). This work was supported by NIH grants R01GM113243, P01AG020677, R01HL104607, G12MD007592 and RR024153, NSF grants DMS1506882, 2G12MD007592 and DMS1512188, and NSA grant H98230-12-1-0246. The authors thank the Editor, Associate Editor and two referees for comments and insights that greatly improved the manuscript.

# Abstract

This article considers the problem of analyzing associations between power spectra of multiple time series and cross-sectional outcomes when data are observed from multiple subjects. The motivating application comes from sleep medicine, where researchers are able to non-invasively record physiological time series signals during sleep. The frequency patterns of these signals, which can be quantified through the power spectrum, contain interpretable information about biological processes. An important problem in sleep research is drawing connections between power spectra of time series signals and clinical characteristics; these connections are key to understanding biological pathways through which sleep affects, and can be treated to improve, health. Such analyses are challenging as they must overcome the complicated structure of a power spectrum from multiple time series as a complex positive-definite matrix-valued function. This article proposes a new approach to such analyses based on a tensor-product spline model of Cholesky components of outcome-dependent power spectra. The approach flexibly models power spectra as nonparametric functions of frequency and outcome while preserving geometric constraints. Formulated in a fully Bayesian framework, a Whittle likelihood based Markov chain Monte Carlo (MCMC) algorithm is developed for automated model fitting and for conducting inference on associations between outcomes and spectral measures. The method is used to analyze data from a study of sleep in older adults and uncovers new insights into how stress and arousal are connected to the amount of time one spends in bed. Supplementary material for this article are available online.

KEY WORDS: Bayesian Analysis; Coherence; Heart Rate Variability; MCMC; Multivariate Time Series; Sleep; Smoothing Spline; Spectral Analysis; Tensor-Product ANOVA; Whittle Likelihood.

## 1 Introduction

Innovations in data collection and storage have led to an increase in the number of biomedical studies that record multiple time series signals and outcome measures in multiple subjects.

For many time series, including common signals such as blood pressure, heart rate and electroencephalography (EEG), frequency patterns that are quantified through the power spectrum contain important information about biological processes. Consequently, studies whose goals are to understand how underlying biological mechanisms are connected to behavioral and clinical outcomes often require an analysis of associations between outcomes and power spectra of multiple time series.

Our motivating application comes from a sleep study whose goal is to better understand the pathways that connect sleep to health and functioning. In the study, heart rate variability (HRV) is recorded in subjects during a night of sleep. HRV is measured through the series of elapsed times between consecutive heart beats, and its power spectrum provides indirect measures of psychological stress and physiological arousal (Hall et al., 2007). Upon awakening, subjects reported subjectively assessed sleep outcomes, such as the amount of time slept during the night, which are associated with many aspects of well-being (Buysse, 2014). Understanding the association between the power spectrum of HRV during different sleep periods (i.e. beginning, middle and end of the night) and self-reported sleep outcomes is essential to understanding how stress connects sleep to health and, consequently, for guiding the use of treatments of poor sleep.

In the biomedical literature, a two-stage approach is typically used to analyze such data. In the first stage, power collapsed within pre-selected frequency bands is estimated individually for each time series (Malik et al., 1996; Hall et al., 2004). A power spectrum is a function of frequency; power collapsed within a frequency band is an integral of the power spectrum over a range of frequencies, which converts the functional parameter into a scalar measure. In the second stage, classical statistical methods, such as ANOVA and linear regression, are used to evaluate associations between these band-collapsed spectral measures and outcomes. Such an approach has three major drawbacks. First, it is highly dependent on the frequency-band collapsed measures selected and there exists a hot debate as to which measures should be considered and/or how they should be interpreted (Burr, 2007). Ideally, an analysis of such data should provide global measures that can be used to understand the entire system

while also providing a means to conduct inference on any frequency band-collapsed measure of potential interest. Second, estimated power is treated as if it were not an estimate but the true unknown parameter, leading to inaccurate inference. Finally, band-specific frequency measures are estimated for each time series separately, inhibiting the evaluation of patterns across series. For instance, in our motivating example, this two-stage approach does not examine how the coherence in HRV between the beginning and end of the night is connected to sleep outcomes.

In the statistics literature, a considerable amount of research has been conducted on methods for analyzing functional variables, a thorough review of which is given by Wang et al. (2016). Included in this body of work are methods for analyzing associations between power spectra and outcomes when one time series is observed per subject (Stoffer et al., 2010; Krafty and Hall, 2013). When one observes multiple time series per subject and interest lies in frequency patterns both within each series and across different series, the problem becomes considerably more challenging. This is the case in our motivating study, where we are interested not only in stress and arousal during particular periods of sleep, but also in their persistence and coherence across periods. While the power spectrum from a single time series is a positive real-valued function of frequency, the power spectrum from multiple time series is a positive-definite Hermitian matrix valued function of frequency. An analysis of associations between power spectra from multiple time series and study outcomes must be able to flexibly model associations while preserving this positive-definite Hermitian structure.

Efficient nonparametric methods that preserve the positive-definite Hermitian structure of spectral matrices have been developed for the simpler, classical problem of estimating the power spectrum of a multivariate time series from a single subject by modeling Cholesky components of spectral matrices as functions of frequency (Dai and Guo, 2004; Rosen and Stoffer, 2007; Krafty and Collinge, 2013). In this article, we extend this framework to develop a new approach to analyzing data from multiple subjects that models Cholesky components as functions of both frequency and outcome. Rather than being curves as functions of

frequency, components of spectral matrices under the proposed model are surfaces. Changes in these surfaces with respect to the outcome provide nonparametric measures of association between outcomes and power spectra. The proposed method is formulated in a fully Bayesian framework; a MCMC algorithm based on the Whittle likelihood, or the asymptotic likelihood derived from the Fourier transform of the data, is developed for model fitting and inference. The method allows one to evaluate the entire outcome-dependent power spectrum and to conduct nonparametric inference on the association between the outcome and any function of the power spectrum.

The rest of the article is organized as follows. Our motivating application, the AgeWise Sleep Study, is discussed in Section 2. A review of spectral analysis in the classical setting, where data are observed from a single subject, is given in Section 3. The proposed methodology for analyzing time series from multiple subjects is presented in Section 4. The proposed method is used to analyze data from the motivating application in Section 5 and some final remarks are offered in Section 6.

## 2 The AgeWise Sleep Study

An estimated 43% of older adults report problems initiating or maintaining sleep (Foley et al., 1995). Poor sleep in older adults has been linked to depression, heart disease, obesity, arthritis, diabetes and stroke (Foley et al., 2004). With medical and scientific advances leading to an increase in the world’s elderly population, the consequences of poor sleep in older adults pose a major public health concern. The AgeWise study is a NIH-funded Program Project conducted at the University of Pittsburgh that seeks a better understanding of causes, effects, and treatments of poor sleep in older adults. Towards this goal, we consider  $N = 108$  men and women between 69–89 years of age who were observed during a night of in-home sleep. Two types of data were collected in each subject. First, subjects were observed during the night through ambulatory polysomnography (PSG), or the continuous collection of electrophysiological changes that occur during sleep. Second, upon awakening, subjects completed the Pittsburgh Sleep Diary (Monk et al., 1994) to record self-reported

sleep outcomes during the night.

As previously discussed, HRV is the series of elapsed times between heart beats. It is of interest to researchers, as it reflects neurological control of the heart, and through this capacity, its power spectrum provides indirect measures of stress reactivity and arousal. The PSG used in the study included an electrocardiograph (ECG) to monitor heart activity. The ECG was used to locate the timing of heart beats, which were then differenced, detrended, cubic spline interpolated, and resampled at 1 Hz to compute HRV series throughout the night.

During the night, the body cycles through two types of sleep: rapid eye movement (REM) and non-rapid eye movement (NREM) sleep. In NREM sleep, which contains deep-sleep, the parasympathetic branch of the autonomic nervous system that is responsible for unconscious actions and stimulates the body to “rest-and-digest” dominates the sympathetic branch, which drives the “flight-or-fight” response. Parasympathetic nervous system activity during NREM is hypothesized to be responsible for many of the rejuvenating properties of sleep (Siegel, 2005). However, physiological activity during NREM sleep is not constant; in good sleepers, the amount of parasympathetic activity during NREM increases throughout the night (Hall et al., 2004). To enable an analysis that can evaluate autonomic nervous system activity and its changes during the night, we consider 3 HRV time series per subject (at the beginning, middle, and end of the night) by extracting the first 5 minutes of HRV from the first three periods of NREM sleep. Data from two subjects are displayed in Figure 1.

The goal of our analysis is to understand how the power spectrum of HRV over the three periods of NREM are connected to self-reported sleep. We focus on one particular self-reported sleep measure derived from the Pittsburgh Sleep Diary: time in bed (TIB). TIB is defined as the elapsed time between attempted sleep and final wakening. It serves as an upper bound for the amount of time spent asleep during the night, which has been linked to heart disease, hypertension, impaired neurobehavioral performance and mortality (Buysse, 2014). The reported TIB from our sample has a mean of 477.99 minutes and a

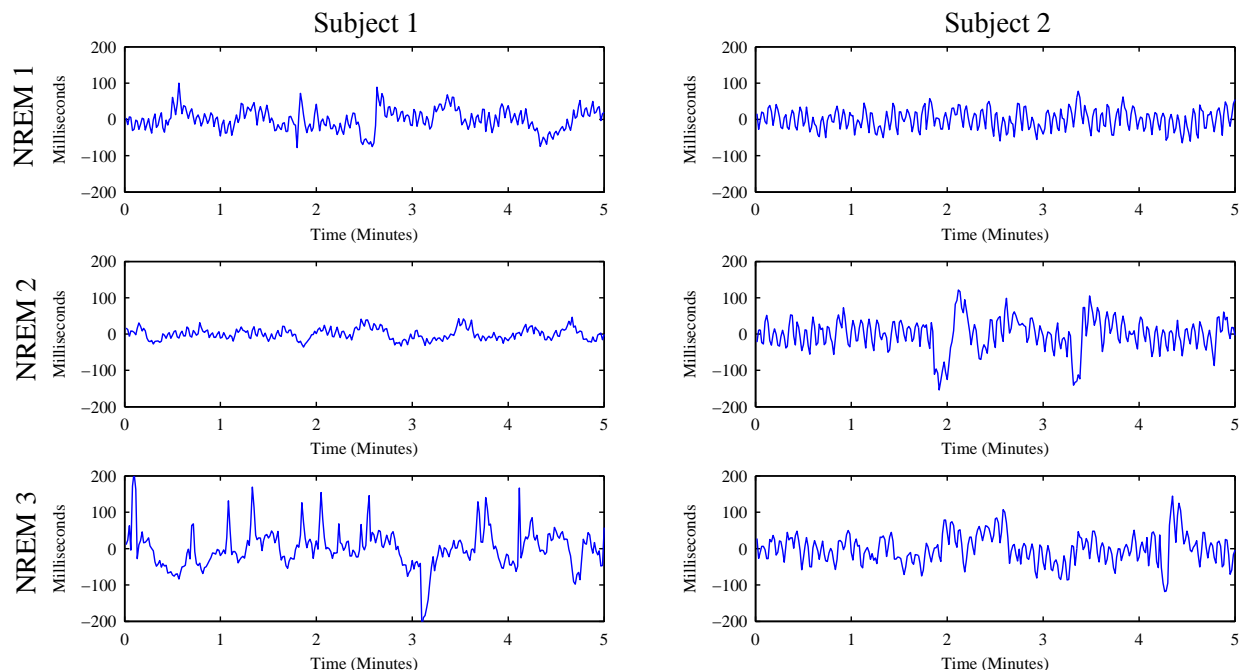


Figure 1: Detrended HRV time series during the first three periods of NREM from two subjects. Subject 1 reported a TIB of 357.67 minutes and subject 2 reported a TIB of 521.00 minutes.

standard deviation of 71.32 minutes. The resulting data for analysis consist of three epochs of HRV, one during each of the first three periods of NREM sleep, and self-reported TIB from each subject.

### 3 Methodological Background: Spectral Domain Analysis

Before introducing our proposed model for the spectral analysis of multiple time series from multiple subjects in Section 4, in this section we present background on spectral analysis in the classical setting, where data are observed from a single subject, for both univariate and multivariate time series.

## 3.1 Univariate Time Series

### 3.1.1 Population Parameters

Spectral domain analysis focuses on the cyclic behavior of time series data. An alternate approach is time domain analysis wherein the relationship between the data at different time lags is the focus. For stationary time series, the main time domain tool is the covariance between a current value of the series, say  $X_t$ , and the value of the series  $h$  time units prior, say  $X_{t-h}$ . The autocovariance is a function of lag, and is given by

$$\gamma(h) = \text{Cov}(X_t, X_{t-h}), \quad h = 0, \pm 1, \pm 2, \dots$$

If  $\gamma(h)$  is absolutely summable (which it is for ARMA models, for example), then there is a duality between the power spectrum, given by

$$f(\omega) = \sum_{h=-\infty}^{\infty} \gamma(h) \exp(-2\pi i \omega h), \quad \omega \in \mathbb{R},$$

and the autocovariance function, namely,

$$\gamma(h) = \int_{-1/2}^{1/2} f(\omega) \exp(2\pi i \omega h) d\omega, \quad h = 0, \pm 1, \pm 2, \dots, \quad (1)$$

as the inverse transform of the power spectrum. The relationship is the same as that of a characteristic function and a probability density. Consequently, the information contained in the power spectrum is equivalent to the information contained in the autocovariance function. If we are concerned with lagged behavior, working with  $\gamma(h)$  is more informative; if we are concerned with cyclic behavior, as is the case of HRV where cyclical behavior provides interpretable physiological information, working with  $f(\omega)$  is more informative.

The power spectrum is nonnegative and we assume that it is positive, so that  $f(\omega) > 0$  for all  $\omega$ . In addition to being positive,  $f$  has two other restrictions as a function of frequency. By the nature of the Fourier transform, it is periodic such that  $f(\omega) = f(\omega + 2\pi)$ , and it is a Hermitian function, or an even function, where  $f(\omega) = f(-\omega)$ . Consequently,  $f(\omega)$  is usually displayed only for  $\omega \in [0, 1/2]$ .



Putting  $h = 0$  in (1) yields

$$\gamma(0) = \text{Var}(X_t) = \int_{-1/2}^{1/2} f(\omega) d\omega,$$

which expresses the total variance of the time series as the integrated power spectrum. In particular, we may think of  $f(\omega) d\omega$  as the approximate variance in the data attributed to frequencies in a small band of width  $d\omega$  around  $\omega$ . It is common to view spectral analysis as an analysis of variance (ANOVA) of time series data with respect to frequency. In fact, the power spectrum is a density of variance rather than of probability.

### 3.1.2 Estimation

The nonparametric estimation of  $f$  from an epoch of length  $n$ ,  $X_1, \dots, X_n$ , can begin by considering the discrete Fourier transform (DFT)

$$Y_m = n^{-1/2} \sum_{t=1}^n X_t \exp(-2\pi i \omega_m t),$$

where  $\omega_m = m/n$  are the Fourier frequencies. When  $n$  is large,  $Y_m$  are approximately independent mean-zero complex normal random variables with variances  $f(\omega_m)$  for  $m = 1, \dots, M$ ,  $M = \lfloor (n-1)/2 \rfloor$ , and  $Y_m = \bar{Y}_{n-m}$  (Shumway and Stoffer, 2011, Appendix C). Consequently, the periodogram  $|Y_m|^2$  provides approximately unbiased but noisy estimates of  $f(\omega_m)$ . Consistent estimates can be obtained by smoothing the periodogram across frequency using tools such as local averaging (Shumway and Stoffer, 2011, Chapter 4.5), splines (Pawitan and O'Sullivan, 1994), and wavelets (Moulin, 1994).

Our estimation approach for multiple time series from multiple subjects, which we develop in Section 4, is based on Bayesian splines. To motivate the development of the new methodology, in this subsection we discuss first a Bayesian smoothing spline model for univariate spectral analysis, then discuss a low-rank approximation. Smoothing spline estimation balances the fit of a function to observed data with a roughness-based measure of regularity. The Bayesian formulation of smoothing splines was first discussed in the case of Gaussian observations by Kimeldorf and Wahba (1970) and Wahba (1978). Under the Bayesian formulation, the likelihood provides a measure of fit to observed data, and regularity is imposed

through a mean-zero Gaussian prior on the functional parameter, which induces a prior for the roughness of the function.

For spectrum estimation, the large sample distribution of  $Y_m$  provides the Whittle likelihood (Whittle, 1953, 1954)

$$L(Y | f) \approx \prod_{m=1}^M f^{-1}(\omega_m) \exp\{-f^{-1}(\omega_m) |Y_m|^2\}.$$

We adopt generic notation throughout this article where  $Y$  will denote all DFT data. Since  $f$  is a positive function,  $\log f$  is modeled rather than  $f$  itself to avoid constraints. Although general measures of regularity can be considered, we focus on measuring the roughness of a function through its integrated squared second derivative,

$$\mathcal{P}(\log f) = \int_0^{1/2} \{[\log f]''(\omega)\}^2 d\omega.$$

The specification of the prior distribution begins by decomposing  $\log f$  into a linear part (which is in the null space of  $\mathcal{P}$ ) and a nonlinear part. To define the prior distribution for the nonlinear part, consider the reproducing kernel of the seminorm defined by  $\mathcal{P}$

$$J(\omega_i, \omega_j) = \int_0^{1/2} (\omega_i - \nu)_+ (\omega_j - \nu)_+ d\nu,$$

where  $(\nu)_+ = \max(\nu, 0)$  (Gu, 2013, Section 2.3.1). For the Bayesian smoothing spline model, the prior distribution for the log-spectrum can be formulated as

$$\log f(\omega) = a_1 + a_2\omega + \sum_{j=1}^M J(\omega, \omega_j)z_j$$

where  $\mathbf{z} = (z_1, \dots, z_M)' \sim N(\mathbf{0}, \tau^2 J^{-1})$  is independent of  $\mathbf{a} = (a_1, a_2)' \sim N(0, \sigma_a^2 I_2)$ ,  $J = \{J(\omega_i, \omega_j)\}$  is the  $M \times M$  matrix of  $J$  evaluated at the Fourier frequencies and  $I_2$  is the  $2 \times 2$  identity matrix. The reproducing property of the kernel  $J$  provides a simple form for the roughness of the log-spectrum,  $\mathcal{P}(\log f) = \mathbf{z}' J \mathbf{z}$  (Gu, 2013, Chapter 2), from which it can be seen that the prior distribution on the coefficients  $\mathbf{z}$  induces a prior on the roughness of the spectrum where  $\mathcal{P}(\log f) \sim \tau^2 \chi_M^2$ , and  $\chi_M^2$  denotes a chi-squared random variable on  $M$  degrees of freedom. The smoothing parameter  $\tau^2 > 0$  balances the smoothness of the

estimator to its fit to the data such that, conditional on  $\tau^2$ , the Bayes estimate (posterior median) approaches a linear function as  $\tau^2 \rightarrow 0$  and interpolates  $|Y_m|^2$  as  $\tau^2 \rightarrow \infty$ .

Two approaches may be taken for estimation and inference under the Bayesian model: empirical Bayes and fully Bayes. In the empirical Bayes approach, a data driven method, such as generalized cross-validation (GCV) or generalized maximum likelihood (GML), is used to select the smoothing parameter  $\tau^2$ . The log-spectrum is then estimated from the posterior distribution conditional on  $\tau^2$ , and its median is equivalent to the frequentist smoothing spline obtained by minimizing a penalized Whittle likelihood as  $\sigma_\alpha^2 \rightarrow \infty$  (Gu, 1992; Qin and Wang, 2008). In the fully Bayesian approach,  $\tau^2$  is treated as a random variable with a noninformative prior, and inference is conducted averaged over the posterior distribution of  $\tau^2$  (Speckman and Sun, 2003; Crainiceanu et al., 2005). Our proposed methodology will adopt the later approach, as discussed in Section 4.

The presented smoothing spline model contains a large number of coefficients, which can impede computation and limit practicality. Low rank approximations, such as those obtained by using a subset of the kernel functions  $J(\cdot, \omega_j)$  (Gu and Kim, 2002) or another set of basis functions contained in the column space of  $J$  (Wood, 2006), can be used to ease computational burden without sacrificing model fit. Here, we consider the basis formed from the scaled eigenvectors of  $J$ , which has been used for power spectrum estimation by Wood et al. (2002), Rosen et al. (2009) and Rosen et al. (2012). This basis can model smooth functions with a relatively few number of basis functions, provides a diagonal prior covariance structure that aids computation, and maintains the intuitive interpretation of the prior distribution regularizing roughness as measured through  $\mathcal{P}$  (Nychka and Cummins, 1996). To formulate this low-rank approximation, we first consider an equivalent formulation of the Bayesian smoothing spline model for  $\log f$  at the Fourier frequencies. Let  $\log \mathbf{f} = [\log f(\omega_1), \dots, \log f(\omega_M)]'$  be the log-spectrum evaluated at the Fourier frequencies. Further, let  $J = V_J D_J V_J'$  be the spectral decomposition of  $J$ ,  $Q_J = V_J D_J^{1/2}$  and  $\mathbf{c} = D_J^{1/2} V_J' \mathbf{z}$ . Then an equivalent model at the Fourier frequencies is

$$\log \mathbf{f} = L_J \mathbf{a} + Q_J \mathbf{c},$$

where  $L_J = \left( \mathbf{1}_M \mid \boldsymbol{\omega} \right)$ ,  $\mathbf{1}_M$  is the  $M$ -vector of ones,  $\boldsymbol{\omega} = (\omega_1, \dots, \omega_M)'$ , and  $\mathbf{c} \sim N(\mathbf{0}, \tau^2 I_M)$ .

The eigenvectors contained in the columns of  $V_J$  are in increasing order of roughness and the eigenvalues contained in the diagonal matrix  $D_J$  decay rapidly (Eubank, 1999). Smooth functions can be accurately modeled through the first several  $n_J < M$  to provide a low-rank approximation. With a slight abuse of notation to avoid the need to introduce further variables, we let  $Q_J$  represent the  $M \times n_J$  matrix of the first  $n_J$  eigenvectors with corresponding coefficients  $\mathbf{c} \sim N(\mathbf{0}, \tau^2 I_{n_J})$ .

The selection of  $n_J$  provides a compromise between low-rank computational feasibility and loss of flexibility relative to full  $M$ -rank model. An intuitive measure of the loss of flexibility is the fraction of the total variance of the covariance matrix  $J$  explained by its first  $n_J$  eigenvectors (FVE), or the sum of its first  $n_J$  eigenvalues divided by the total sum of all of its  $M$  eigenvalues. Wood et al. (2002), Rosen et al. (2009) and Rosen et al. (2012) suggest using  $n_J = 10$  basis functions, which, in each of their settings considered, equates to an FVE of 97.975%. Our empirical findings support this suggestion, and we recommend selecting  $n_J$  to achieve a 97.975% FVE. Under this rule,  $n_J = 7$  for  $n = [15, 18]$ ,  $n_J = 8$  for  $n = [19, 22]$ ,  $n_J = 9$  for  $n \in [23, 40]$  and  $n_J = 10$  for  $n \in [41, 10^4]$ .

## 3.2 Multivariate Time Series

### 3.2.1 Population Parameters

The ideas presented in the univariate case generalize to the multivariate case wherein we observe a  $P$ -dimensional vector-valued time series, say  $\{\mathbf{X}_t\}$ . Under stationarity, the auto-covariance function is a  $P \times P$  matrix given by

$$\Gamma(h) = \text{Cov}(\mathbf{X}_t, \mathbf{X}_{t-h}), \quad h = 0, \pm 1, \pm 2, \dots$$

If  $\sum_h \|\Gamma(h)\| < \infty$ , the spectral density matrix of the series  $\mathbf{X}_t$  is given by

$$f(\boldsymbol{\omega}) = \sum_{h=-\infty}^{\infty} \Gamma(h) \exp(-2\pi i \boldsymbol{\omega} h), \quad \boldsymbol{\omega} \in \mathbb{R}$$

and the inverse relationship is

$$\Gamma(h) = \int_{-1/2}^{1/2} f(\omega) \exp(2\pi i \omega h) d\omega, \quad h = 0, \pm 1, \pm 2, \dots$$

For each  $\omega \in \mathbb{R}$ ,  $f(\omega)$  is a  $P \times P$  non-negative definite Hermitian matrix with the diagonal elements,  $f_{pp}(\omega)$  for  $p = 1, \dots, P$ , being the spectra of the individual components, and the off-diagonal elements,  $f_{qp}(\omega)$  for  $q \neq p = 1, \dots, P$ , being the cross-spectra. Throughout this article, we assume that  $f(\omega)$  is non-singular for all  $\omega \in \mathbb{R}$ . As in the univariate case,  $f$  is a periodic and Hermitian function of frequency were, for matrix-valued functions, Hermitian as a function of frequency is defined as  $f(\omega) = f^*(-\omega)$ , and  $f^*(\omega)$  is the complex conjugate of  $f(\omega)$ .

An important example of the application of the cross-spectrum is to the problem of linearly predicting one of the component series, say  $X_{qt}$ , from another component, say  $X_{pt}$ . A measure of the strength of such a relationship is the squared coherence function defined as

$$\rho_{qp}^2(\omega) = \frac{|f_{qp}(\omega)|^2}{f_{qq}(\omega)f_{pp}(\omega)}.$$

This is analogous to conventional squared correlation between two finite-variance random variables; e.g.,  $0 \leq \rho_{qp}^2(\omega) \leq 1$ . This analogy motivates the interpretation of squared coherence as the squared correlation between two time series at frequency  $\omega$ . These ideas extend in an obvious way to the concept of multiple coherence and partial coherence functions obtained from the full spectral matrix in much the same way that multiple correlation and partial correlation can be obtained from a covariance matrix. Full details of these results may be found in Shumway and Stoffer (2011, Chapters 4 & 7).

### 3.2.2 Estimation

In the multivariate setting, let

$$\mathbf{Y}_m = n^{-1/2} \sum_{t=1}^n \mathbf{X}_t \exp(-2\pi i \omega_m t)$$

be the DFTs of the data. In this case, the Whittle likelihood is

$$L(Y | f) \approx \prod_{m=1}^M |f^{-1}(\omega_m)| \exp\{-\mathbf{Y}_m^* f^{-1}(\omega_m) \mathbf{Y}_m\},$$

and the periodogram  $\mathbf{Y}_m \mathbf{Y}_m^*$  is an approximately unbiased but noisy estimate of  $f(\omega_m)$ , from which consistent estimates can be obtained by smoothing.

While in the univariate setting the spectrum is smoothed on the logarithmic scale to preserve positivity, Cholesky components of spectral matrices can be smoothed to preserve positive-definiteness in the multivariate setting (Dai and Guo, 2004; Rosen and Stoffer, 2007; Krafty and Collinge, 2013). The modified Cholesky decomposition assures that, for a spectral matrix  $f(\omega)$ , there exists a unique  $P \times P$  lower triangular complex matrix  $\Theta(\omega)$  with ones on the diagonal and a unique  $P \times P$  positive diagonal matrix  $\Psi(\omega)$  such that

$$f^{-1}(\omega) = \Theta(\omega) \Psi^{-1}(\omega) \Theta^*(\omega).$$

There are  $P^2$ -Cholesky components to estimate:  $\Re\{\Theta_{k\ell}\}$  and  $\Im\{\Theta_{k\ell}\}$  for  $k > \ell = 1, \dots, P-1$ , and  $\Psi_{kk}^{-1}$  for  $k = 1, \dots, P$ . Since the diagonal terms  $\Psi_{kk}^{-1}(\omega) > 0$ , we model  $\log \Psi_{kk}^{-1}$ . Letting  $\boldsymbol{\theta}_{k\ell} = [\Theta_{k\ell}(\omega_1), \dots, \Theta_{k\ell}(\omega_M)]'$  and  $\log \boldsymbol{\psi}_{kk}^{-1} = [\log \Psi_{kk}^{-1}(\omega_1), \dots, \log \Psi_{kk}^{-1}(\omega_M)]'$ , we model:

$$\Re\{\boldsymbol{\theta}_{k\ell}\} = L_J \mathbf{a}_{rkl} + Q_J \mathbf{c}_{rkl}, \quad k > \ell = 1, \dots, P-1 \quad (2)$$

$$\Im\{\boldsymbol{\theta}_{k\ell}\} = L_J \mathbf{a}_{ikl} + Q_J \mathbf{c}_{ikl}, \quad k > \ell = 1, \dots, P-1 \quad (3)$$

$$\log \boldsymbol{\psi}_{kk}^{-1} = L_J \mathbf{a}_{dkk} + Q_J \mathbf{c}_{dkk}, \quad k = 1, \dots, P, \quad (4)$$

where  $\mathbf{c}_{rkl} \sim N(\mathbf{0}, \tau_{rkl}^2 I_n)$ ,  $\mathbf{c}_{ikl} \sim N(\mathbf{0}, \tau_{ikl}^2 I_n)$ ,  $\mathbf{c}_{dkk} \sim N(\mathbf{0}, \tau_{dkk}^2 I_n)$ ,  $\mathbf{a}_{rkl} \sim N(\mathbf{0}, \sigma_\alpha^2 I_2)$ ,  $\mathbf{a}_{ikl} \sim N(\mathbf{0}, \sigma_\alpha^2 I_2)$  and  $\mathbf{a}_{dkk} \sim N(\mathbf{0}, \sigma_\alpha^2 I_2)$ . Throughout this article,  $r$ ,  $i$  and  $d$  are used to denote coefficients for real components of  $\Theta$ , imaginary components of  $\Theta$  and the logarithm of the diagonal components of  $\Psi^{-1}$ , respectively.

## 4 Methodology: Replicated Multiple Time Series

The primary question considered in this article is how to assess the association between the power spectrum of  $P$ -variate time series of length  $n$ ,  $\{\mathbf{X}_{j1}, \dots, \mathbf{X}_{jn}\}$ , and real-valued static

variables,  $U_j$ , observed from  $j = 1, \dots, N$  independent subjects. In the motivating study, there are  $N = 108$  participants,  $U_j$  is self-reported TIB, and  $\mathbf{X}_{jt}$  are time series of HRV during the first  $P = 3$  periods of NREM. To address this question, we first introduce a new measure in Section 4.1, the conditional power spectrum, which quantifies associations between power spectra and outcomes. Then, in Section 4.2, we develop a tensor product model for the conditional power spectrum that extends the Bayesian spline model of Cholesky components of a single multivariate time series to account for dependence on both frequency and outcome.

As previously mentioned, there are two approaches to conducting a Bayesian analysis with splines: empirical Bayes and fully Bayesian. Each approach has strengths and weaknesses. In the empirical Bayes approach, smoothing parameters are estimated through a data-driven procedure. Estimates conditional on smoothing parameters can be quickly computed through Fisher’s scoring or Newton-Raphson and conditional inference on the modeled functions (Cholesky components in our setting) can be conducted through approximate “Bayesian confidence intervals” (Gu, 1992). In the fully Bayesian approach, smoothing parameters  $\tau^2$  are treated as random variables with noninformative priors and MCMC techniques are used to sample from the joint distribution of coefficients and smoothing parameters (Speckman and Sun, 2003; Crainiceanu et al., 2005). The sample simulated from the posterior distribution using MCMC provides a natural means of conducting inference on any function of the spectrum averaged over the distribution of the smoothing parameters, which accounts for uncertainty in the smoothing parameters when conducting inference. As will be illustrated in Section 5, inference on squared coherence, univariate spectra, and integral functions thereof (all of which are nonlinear functions of modeled Cholesky components) are of direct scientific interest. We develop the proposed methodology under a fully Bayesian framework, presenting prior distributions in Section 4.3 and the sampling scheme in Section 4.4.

## 4.1 Conditional Power Spectrum

Without loss of generality, we formulate the methodology assuming that  $U_i$  is scaled to take values within  $[0, 1]$ . To quantify the association between the power spectrum of the time series  $\mathbf{X}_{jt}$  and the outcome  $U_j$ , we define the conditional power spectrum

$$f(\omega, u) = \sum_{h=-\infty}^{\infty} \text{Cov}(\mathbf{X}_{jt}, \mathbf{X}_{j,t+h} \mid U_j = u) e^{-2\pi i \omega h}, \quad \omega \in \mathbb{R}, u \in [0, 1].$$

As with the power spectrum of a single multivariate time series, the spectral matrices  $f(\omega, u)$  are positive-definite  $P \times P$  Hermitian matrices, and  $f(\cdot, u)$  is a periodic and Hermitian function of frequency for fixed  $u$ . In a traditional spectral analysis without a cross-sectional variable, spectral measures such as  $f_{pq}$  and  $\rho_{pq}^2 = |f_{pq}|^2 / (f_{pp}f_{qq})$  are curves as functions of frequency. In the conditional setting, these are surfaces as functions of both frequency and the variable  $u$ . How these functions change with respect to  $u$  provides information as to how spectral measures are associated with the variable.

## 4.2 Bayesian Tensor-Product Model of Cholesky Components

As in the classical setting discussed in Section 3.2.2, where a multivariate time series observed from a single subject, to preserve positive definiteness, we model the Cholesky components. Let

$$f^{-1}(\omega, u) = \Theta(\omega, u)\Psi^{-1}(\omega, u)\Theta^*(\omega, u),$$

be the modified Cholesky decomposition of the conditional power spectrum. We use Bayesian tensor product models for the  $P^2$ -unique Cholesky components, which decompose the bivariate functions into products of univariate functions of  $\omega$  and of  $u$ .

Bayesian models for Cholesky components as functions of  $\omega$  were discussed in Section 3.2.2. Similarly, a low-rank approximate Bayesian smoothing spline model for a function of outcomes at the observed values can be formulated. Since the domain of the outcome values is  $[0, 1]$ , as opposed the domain of the frequency values  $[0, 1/2]$ , we consider the kernel

$$H(u_i, u_j) = \int_0^1 (u_i - v)_+ (u_j - v)_+ dv.$$



Letting  $\mathbf{u} = (u_1, \dots, u_N)'$ , a low-rank model for functions of the outcome evaluated at the observed values is:

$$L_H \mathbf{a} + Q_H \mathbf{b}, \quad (5)$$

where  $L_H = \left( \mathbf{1}_N \mid \mathbf{u} \right)$ ,  $Q_H$  is the  $N \times n_H$  matrix of the first  $n_H$  columns of  $V_H D_H^{1/2}$ ,  $H = V_H D_H^{1/2} V_H'$  is the spectral decomposition of the  $N \times N$  matrix  $H = \{H(u_i, u_j)\}$ , and  $\mathbf{a} \sim N(\mathbf{0}, \sigma_\alpha^2 I_2)$  is independent of  $\mathbf{b} \sim N(\mathbf{0}, \tau^2 I_{n_H})$ .

To write the tensor-product model at the observed frequency-outcome points, concatenate components across frequency and outcome to define the  $NM$ -vectors

$$\boldsymbol{\theta}_{k\ell} = [\{\Theta_{k\ell}(\omega_1, u_1), \dots, \Theta_{k\ell}(\omega_M, u_1)\}, \dots, \{\Theta_{k\ell}(\omega_1, u_N), \dots, \Theta_{k\ell}(\omega_M, u_N)\}]'$$

for  $k > \ell = 1, \dots, P-1$ . Similarly define  $\log \boldsymbol{\psi}_{kk}^{-1}$  for  $k = 1, \dots, P$ . The real and imaginary parts of  $\boldsymbol{\theta}_{k\ell}$ , and  $\log \boldsymbol{\psi}_{kk}^{-1}$  can then be expressed as tensor products of the spline models for functions of frequency (given in Equations (2) - (4)) and outcome (given in Equation (5))

$$\begin{aligned} \Re\{\boldsymbol{\theta}_{k\ell}\} &= \{L_H \otimes L_J\} \mathbf{a}_{rkl} + \{Q_H \otimes L_J\} \mathbf{b}_{rkl} + \{L_H \otimes Q_J\} \mathbf{c}_{rkl} + \{Q_H \otimes Q_J\} \mathbf{d}_{rkl} \\ \Im\{\boldsymbol{\theta}_{k\ell}\} &= \{L_H \otimes L_J\} \mathbf{a}_{ikl} + \{Q_H \otimes L_J\} \mathbf{b}_{ikl} + \{L_H \otimes Q_J\} \mathbf{c}_{ikl} + \{Q_H \otimes Q_J\} \mathbf{d}_{ikl} \\ \log \boldsymbol{\psi}_{kk}^{-1} &= \{L_H \otimes L_J\} \mathbf{a}_{dkk} + \{Q_H \otimes L_J\} \mathbf{b}_{dkk} + \{L_H \otimes Q_J\} \mathbf{c}_{dkk} + \{Q_H \otimes Q_J\} \mathbf{d}_{dkk}. \end{aligned}$$

This model decomposes conditional Cholesky components into combinations of products of univariate functions of frequency and univariate functions of outcome. The parameters  $\mathbf{a}$  are coefficients for functions that are products of linear functions of both  $\omega$  and  $u$ ,  $\mathbf{b}$  are coefficients for functions that are products of linear functions of  $\omega$  and nonlinear functions of  $u$ ,  $\mathbf{c}$  are coefficients for functions that are products of nonlinear functions of  $\omega$  and linear functions of  $u$ , and  $\mathbf{d}$  are coefficients for functions that are products of nonlinear functions of  $\omega$  and of  $u$ .

### 4.3 Prior Distributions

We define two types of prior distributions: prior distributions on coefficients conditional on smoothing parameters and prior distributions on smoothing parameters. The tensor product model naturally enables the formulation of prior distributions that regularize its components

as univariate functions of frequency and outcome. Letting  $\boldsymbol{\eta}_{rkl} = (\mathbf{a}'_{rkl}, \mathbf{b}'_{rkl}, \mathbf{c}'_{rkl}, \mathbf{d}'_{rkl})'$ ,  $\boldsymbol{\eta}_{ikl} = (\mathbf{a}'_{ikl}, \mathbf{b}'_{ikl}, \mathbf{c}'_{ikl}, \mathbf{d}'_{ikl})'$  and  $\boldsymbol{\eta}_{dkk} = (\mathbf{a}'_{dkk}, \mathbf{b}'_{dkk}, \mathbf{c}'_{dkk}, \mathbf{d}'_{dkk})'$ , conditional on smoothing parameters, we assume the diagonal Gaussian smoothing priors

$$\begin{aligned}\boldsymbol{\eta}_{rkl} &\sim N(\mathbf{0}, D_{rkl}) \text{ where } D_{rkl} = \text{diag}(\sigma_\alpha^2 \mathbf{1}'_4, \tau_{\beta rkl}^2 \mathbf{1}'_{2n_H}, \tau_{\gamma rkl}^2 \mathbf{1}'_{2n_J}, \tau_{\delta rkl}^2 \mathbf{1}'_{n_H \times n_J}), \\ \boldsymbol{\eta}_{ikl} &\sim N(\mathbf{0}, D_{ikl}) \text{ where } D_{ikl} = \text{diag}(\sigma_\alpha^2 \mathbf{1}'_4, \tau_{\beta ikl}^2 \mathbf{1}'_{2n_H}, \tau_{\gamma ikl}^2 \mathbf{1}'_{2n_J}, \tau_{\delta ikl}^2 \mathbf{1}'_{n_H \times n_J}), \\ \boldsymbol{\eta}_{dkk} &\sim N(\mathbf{0}, D_{dkk}) \text{ where } D_{dkk} = \text{diag}(\sigma_\alpha^2 \mathbf{1}'_4, \tau_{\beta dkk}^2 \mathbf{1}'_{2n_H}, \tau_{\gamma dkk}^2 \mathbf{1}'_{2n_J}, \tau_{\delta dkk}^2 \mathbf{1}'_{n_H \times n_J})\end{aligned}$$

where  $\mathbf{1}_n$  is the  $n$ -vector of ones.

Prior distributions on the smoothing parameters are placed by assuming that  $\tau_{\beta rkl}, \tau_{\gamma rkl}, \tau_{\delta rkl}, \tau_{\beta ikl}, \tau_{\gamma ikl}, \tau_{\delta ikl}$ ,  $k > \ell = 1, \dots, P-1$ ,  $\tau_{\beta dkk}, \tau_{\gamma dkk}, \tau_{\delta dkk}$ ,  $k = 1, \dots, P$ , are independent Half- $t(\nu, G)$  random variables with pdf  $p(x) \propto [1 + (x/G)^2/\nu]^{-(\nu+1)/2}$ ,  $x > 0$ , where the hyperparameters  $\nu$  and  $G$  are assumed known (Gelman, 2006). Computationally, it is convenient to utilize the following scale mixture representation (Wand et al., 2012):  $(\tau^2 \mid g) \sim IG(\nu/2, \nu/g)$ ,  $g \sim IG(1/2, 1/G^2)$ , where  $IG(a, b)$ , is the inverse Gamma distribution with pdf  $p(x) \propto x^{-(a+1)} \exp(-b/x)$ ,  $x > 0$ . The larger the value of  $G$ , the less informative the prior, and we set  $G$  to a large fixed value. We found analyses to be insensitive to the choice of  $G$ , with  $G = 10$  and  $G = 10^5$  giving indistinguishable results in both simulations and in the analysis of the AgeWise data. The hyperparameter  $\sigma_\alpha^2$ , which is the prior variance of the coefficients of the linear terms, is assumed to be a known large value. In our computations,  $\sigma_\alpha^2 = 10^2$  and  $\sigma_\alpha^2 = 10^7$  gave indistinguishable results.

## 4.4 Whittle Likelihood, Sample Scheme and Inference

Given observed time series, we define the DFT for the  $j$ th subject at frequency  $\omega_m$  as

$$\mathbf{Y}_{jm} = n^{-1/2} \sum_{t=1}^n \mathbf{X}_{jt} \exp(-2\pi i \omega_m t).$$

For large  $n$ , conditional on  $u_j$ ,  $\mathbf{Y}_{jm}$  are approximately independent mean-zero complex Gaussian random variables. This provides the conditional Whittle likelihood

$$L(Y \mid f) \approx \prod_{j=1}^N \prod_{m=1}^M |f^{-1}(\omega_m, u_j)| \exp\{-\mathbf{Y}_{jm}^* f^{-1}(\omega_m, u_j) \mathbf{Y}_{jm}\}.$$

There are  $P^2 [(n_H + 2)(n_J + 2) + 3]$  parameters in the model of  $f$ :  $(n_H + 2)(n_J + 2)$  regression coefficients and 3 smoothing parameters for each of the  $P^2$  Cholesky components. We develop a sampling scheme to sample from the joint posterior distribution of the coefficients  $\boldsymbol{\eta}$ 's and smoothing parameters  $\tau^2$ 's conditional on the DFT  $Y$  and the observed outcomes  $\mathbf{u}$ . To aid in developing this sampling scheme, it is advantageous to consider a more compact notation by defining

$$Q = \left( L_H \otimes L_J \quad \middle| \quad Q_H \otimes L_J \quad \middle| \quad L_H \otimes Q_J \quad \middle| \quad Q_H \otimes Q_J \right)$$

so that

$$\Re\{\boldsymbol{\theta}_{k\ell}\} = Q\boldsymbol{\eta}_{r_{k\ell}}, \quad \Im\{\boldsymbol{\theta}_{k\ell}\} = Q\boldsymbol{\eta}_{i_{k\ell}} \quad \text{and} \quad \log \boldsymbol{\psi}_{kk}^{-1} = Q\boldsymbol{\eta}_{d_{kk}}.$$

Each iteration of the sampling scheme consists of three steps. First, the coefficients corresponding to the real and imaginary components of  $\Theta$  ( $\boldsymbol{\eta}_{r_{k\ell}}$  and  $\boldsymbol{\eta}_{i_{k\ell}}$ ) are sequentially sampled as Gaussian random variables from their conditional posterior distributions conditional on the current values of all other parameters. In the second step, the coefficients corresponding to the diagonal elements of  $\Psi^{-1}$  ( $\boldsymbol{\eta}_{d_{kk}}$ ) are drawn. The log of the conditional posterior distribution of these coefficients is given by

$$\log p(\boldsymbol{\eta}_{d_{kk}} \mid \mathbf{v}_k, D_{d_{kk}}) \stackrel{c}{=} \sum_{j=1}^N \sum_{m=1}^M \left\{ \mathbf{q}'_{jm} \boldsymbol{\eta}_{d_{kk}} - \exp(\mathbf{q}'_{jm} \boldsymbol{\eta}_{d_{kk}}) v_{kjm} \right\} - \frac{1}{2} \boldsymbol{\eta}'_{d_{kk}} D_{d_{kk}}^{-1} \boldsymbol{\eta}_{d_{kk}}, \quad (6)$$

where  $\mathbf{q}'_{jm}$  is the row of  $Q$  corresponding to the  $j$ th subject and  $m$ th frequency,  $\mathbf{v}_k$  is a vector with components  $v_{kjm}$  depending on  $Y$  and on other parameters held fixed (its exact form is given in Appendix B), and  $\stackrel{c}{=}$  denotes equality up to a constant. Since this is not a known distribution,  $\boldsymbol{\eta}_{d_{kk}}$  are drawn in a Metropolis-Hastings step. The last step samples smoothing parameters from their posterior distributions conditional on other parameters. For ease of notation, in what follows we describe the sampling scheme for the case  $P = 3$ . Further details are given in Appendix B of the online supplementary material. After initializing all the parameters, the  $s$ th iteration,  $1 \leq s \leq S$ , of the Gibbs sampler consists of the following steps.

1. Sample the coefficients corresponding to  $\Theta$ :

(a) Draw

$$(\boldsymbol{\eta}_{r21}^{(s)} | Y, \boldsymbol{\psi}_{11}^{-1(s-1)}, \boldsymbol{\theta}_{31}^{(s-1)}, D_{r21}^{(s-1)}) \sim N(\boldsymbol{\mu}_{r21}, \Sigma_{r21})$$

$$(\boldsymbol{\eta}_{i21}^{(s)} | Y, \boldsymbol{\psi}_{11}^{-1(s-1)}, \boldsymbol{\theta}_{31}^{(s-1)}, D_{i21}^{(s-1)}) \sim N(\boldsymbol{\mu}_{i21}, \Sigma_{i21})$$

and update  $\boldsymbol{\theta}_{21}^{(s)} = Q\boldsymbol{\eta}_{r21}^{(s)} + iQ\boldsymbol{\eta}_{i21}^{(s)}$ .

(b) Draw

$$(\boldsymbol{\eta}_{r31}^{(s)} | Y, \boldsymbol{\psi}_{11}^{-1(s-1)}, \boldsymbol{\theta}_{21}^{(s)}, D_{r31}^{(s-1)}) \sim N(\boldsymbol{\mu}_{r31}, \Sigma_{r31})$$

$$(\boldsymbol{\eta}_{i31}^{(s)} | Y, \boldsymbol{\psi}_{11}^{-1(s-1)}, \boldsymbol{\theta}_{21}^{(s)}, D_{i31}^{(s-1)}) \sim N(\boldsymbol{\mu}_{i31}, \Sigma_{i31})$$

and update  $\boldsymbol{\theta}_{31}^{(s)} = Q\boldsymbol{\eta}_{r31}^{(s)} + iQ\boldsymbol{\eta}_{i31}^{(s)}$ .

(c) Draw

$$(\boldsymbol{\eta}_{r32}^{(s)} | Y, \boldsymbol{\psi}_{22}^{-1(s-1)}, D_{r32}^{(s-1)}) \sim N(\boldsymbol{\mu}_{r32}, \Sigma_{r32})$$

$$(\boldsymbol{\eta}_{i32}^{(s)} | Y, \boldsymbol{\psi}_{22}^{-1(s-1)}, D_{i32}^{(s-1)}) \sim N(\boldsymbol{\mu}_{i32}, \Sigma_{i32})$$

and update  $\boldsymbol{\theta}_{32}^{(s)} = Q\boldsymbol{\eta}_{r32}^{(s)} + iQ\boldsymbol{\eta}_{i32}^{(s)}$ .

The exact forms of the conditional means and covariances,  $\mu_{ckl}$  and  $\Sigma_{ckl}$ ,  $c = r, i$ , are given in Appendix B.

2. Sample coefficients corresponding to  $\Psi^{-1}$ :

**for**  $k = 1, 2, 3$  **do**

(a) Draw  $\boldsymbol{\eta}_{dkk}^{(s)} \sim t_\nu(\hat{\boldsymbol{\eta}}_{dkk}, \hat{\Sigma}_{dkk})$ , where  $\hat{\boldsymbol{\eta}}_{dkk}$  is the maximizer of (6) and  $\hat{\Sigma}_{dkk}$  is the inverse of the observed information matrix at  $\hat{\boldsymbol{\eta}}_{dkk}$ .

(b) Compute

$$r^{(s)} = \frac{p(\boldsymbol{\eta}_{dkk}^{(s)} | \mathbf{v}_k, D_{dkk}) f_T(\boldsymbol{\eta}_{dkk}^{(s-1)})}{p(\boldsymbol{\eta}_{dkk}^{(s-1)} | \mathbf{v}_k, D_{dkk}) f_T(\boldsymbol{\eta}_{dkk}^{(s)})},$$

where  $f_T$  is the density of the  $t_\nu(\hat{\boldsymbol{\eta}}_{dkk}, \hat{\Sigma}_{dkk})$  distribution.

(c) With probability  $\min(1, r^{(s)})$  accept  $\boldsymbol{\eta}_{dkk}^{(s)}$ , otherwise  $\boldsymbol{\eta}_{dkk}^{(s)} = \boldsymbol{\eta}_{dkk}^{(s-1)}$ .

(d) Update  $\boldsymbol{\psi}_{kk}^{-1(s)} = \exp(Q\boldsymbol{\eta}_{dkk}^{(s)})$ .

**end**

3. Sample smoothing parameters:

```

for  $\ell = 1, 2$  do
  for  $k = 2, 3$  do
    Draw
     $\tau_{\beta rkl}^{2(s)} \sim IG((n_b + \nu)/2, \mathbf{b}'_{rkl} \mathbf{b}_{rkl}^{(s)}/2 + \nu/g_{\beta rkl}^{(s-1)})$ 
     $\tau_{\gamma rkl}^{2(s)} \sim IG((n_c + \nu)/2, \mathbf{c}'_{rkl} \mathbf{c}_{rkl}^{(s)}/2 + \nu/g_{\gamma rkl}^{(s-1)})$ 
     $\tau_{\delta rkl}^{2(s)} \sim IG((n_d + \nu)/2, \mathbf{d}'_{rkl} \mathbf{d}_{rkl}^{(s)}/2 + \nu/g_{\delta rkl}^{(s-1)})$ 
     $g_{\beta rkl}^{(s)} \sim IG((\nu + 1)/2, \nu/\tau_{\beta rkl}^{2(s)} + 1/G^2)$ 
     $g_{\gamma rkl}^{(s)} \sim IG((\nu + 1)/2, \nu/\tau_{\gamma rkl}^{2(s)} + 1/G^2)$ 
     $g_{\delta rkl}^{(s)} \sim IG((\nu + 1)/2, \nu/\tau_{\delta rkl}^{2(s)} + 1/G^2)$ 
  end
end

```

The smoothing parameters for the imaginary and diagonal components are similarly drawn from inverse gamma distributions.

## Point Estimates and Credible Intervals

The sample generated via MCMC methods provides a means of obtaining point estimates and credible intervals for any function of the spectrum averaged over the distribution of smoothing parameters through the sample mean and percentiles of the empirical distribution of the function evaluated at each iteration of the sampling algorithm. For instance, a measure of interest in the analysis of HRV is the log-spectrum from the  $p$ th period of NREM,  $\log f_{pp}$ . Consider

$$f^{(s)}(\omega_m, u_j) = \{\Theta^{(s)}(\omega_m, u_j) \Psi^{-1(s)}(\omega_m, u_j) \Theta^{*(s)}(\omega_m, u_j)\}^{-1}$$

as the estimated spectral matrix at the  $s$ th iteration corresponding to  $u_j$  and  $\omega_m$  with  $p$ th diagonal element  $f_{pp}^{(s)}(\omega_m, u_j)$ . The matrix  $\Theta^{(s)}(\omega_m, u_j)$  has  $k\ell$ th element

$$\theta_{k\ell jm}^{(s)} = \mathbf{q}'_{jm} \boldsymbol{\eta}_{rkl}^{(s)} + i \mathbf{q}'_{jm} \boldsymbol{\eta}_{ikl}^{(s)}, \quad k > \ell = 1, \dots, P-1,$$

and  $\Psi^{-1(s)}(\omega_m, u_j)$  has  $kk$ th element  $\exp(\mathbf{q}'_{kk} \boldsymbol{\eta}_{dkk}^{(s)})$ . If  $S$  iterations of the sampling algorithm are run with a burn-in of  $S_0$ , then an estimate of  $\log f_{pp}(\omega_m, u_\ell)$  can be computed as the mean of the values  $\{\log f_{pp}^{(s)}(\omega_m, u_\ell); S_0 \leq s \leq S\}$ , and a 95% credible interval computed as their 2.5 and 97.5 empirical percentiles.

Scientific interest also lies in measures collapsed across frequency. For example, as will be discussed in the following section, in the analysis of HRV, collapsed power within the high-frequency band (HF) between 0.15-0.40 Hz

$$f_p^{HF}(u) = \int_{.15}^{.40} f_{pp}(\omega, u) d\omega \quad (7)$$

provides a measure of autonomic nervous system activity during the  $p$ th NREM period among people with a TIB of  $u$ . Letting  $f_p^{HF(s)}(u_\ell) = W^{-1} \sum_{.15 \leq \omega_m \leq .40} f_{pp}^{(s)}(\omega_m, u_\ell)$  where  $W$  is the number of Fourier frequencies within the HF band, an estimate of  $f_p^{HF}(u_\ell)$  is given by the mean of the values  $\{f_p^{HF(s)}(u_\ell); S_0 \leq s \leq S\}$ , and a 95% credible interval is given by their 2.5 and 97.5 empirical percentiles.

## 5 Application to the AgeWise Study

We used the proposed methodology to analyze the association between TIB and the power spectrum of the first three periods of NREM from  $N = 108$  AgeWise subjects, as described in Section 2. The method was fit using  $n_H = n_J = 10$  basis functions, with hyperparameters  $G = \sigma_\alpha^2 = 10^5$ , and for  $S = 3000$  iterations of the MCMC algorithm with a burn-in of  $S_0 = 500$ . Note that, in this example, there are a total of  $P^2(n_H + 2)(n_J + 2) = 1296$  coefficients and  $3P^2 = 27$  smoothing parameters. The average run time per iteration was 5.46 seconds with a standard deviation of 0.24 seconds using the program that is available on the journal's website in Matlab 2016b and macOS Sierra v10.12.1 on a 2.9 GHz Intel Core i7 processor with 16 GB RAM.

Although all desired analyses are obtained from one MCMC chain, to aid the biological and clinical discussion of the results, we present the analysis in three stages. First, in Section 5.1 we examine the estimated period-specific spectra and squared coherences as frequency-outcome surfaces. In the subsequent two stages, we explore power collapsed within certain frequency bands as functions of TIB: first for power within periods in Section 5.2, then for coherence between periods in Section 5.3. The results from these analyses provide new insights into biological underpinnings of spending too little or too much time in bed. In

particular, our analysis suggests that (i) short TIB is connected to elevated stress and arousal within-periods of NREM towards the end of the night and (ii) long TIB is associated with a persistence in arousal in the beginning of the night.

## 5.1 Analysis of the Conditional Spectrum

Point estimates of the within-period conditional log-spectral surfaces,  $\log \{f_{pp}(\omega, u)\}$ , and of the cross-period conditional logit squared coherence surfaces,

$$\text{logit} \{ \rho_{pq}^2(\omega, u) \} = \log \left[ \rho_{pq}^2(\omega, u) / \{1 - \rho_{pq}^2(\omega, u)\} \right],$$

are displayed in Figure 2. These estimates are plotted on the logarithmic and logistic scales, respectively, to aid visualization. The conditional spectra at each of the first three periods of NREM and the squared coherence between NREM 1 and 2 display different characteristics within low frequencies that are less than 0.15 Hz compared to higher frequencies between 0.15–0.40 Hz.

From a biological perspective, these results are not surprising and produce interpretable measures. As was discussed in Section 2, the autonomic nervous system is classically divided into two branches: the parasympathetic branch that is responsible for activities related to resting and digestion and the sympathetic branch that is responsible for the flight-or-flight response. Researchers have shown that power within the high frequency band (HF) within 0.15-0.40 Hz provides a measure of parasympathetic nervous system activity and that power within the low frequency band (LF) between 0.04-0.15 Hz is a measure of the combined modulation of both the sympathetic and parasympathetic nervous systems. Consequently, the ratio of power from low frequencies versus high frequencies (LF/HF) can be interpreted as a measure of sympathetic modulation relative to parasympathetic modulation. Blunted HF and elevated LF/HF power are often interpreted as indirect measures of physiological arousal and psychological stress (Hall et al., 2004, 2007). To obtain inference on associations between these measures and TIB, in the next two subsections we examine power and coherence collapsed within these bands as functions of TIB.

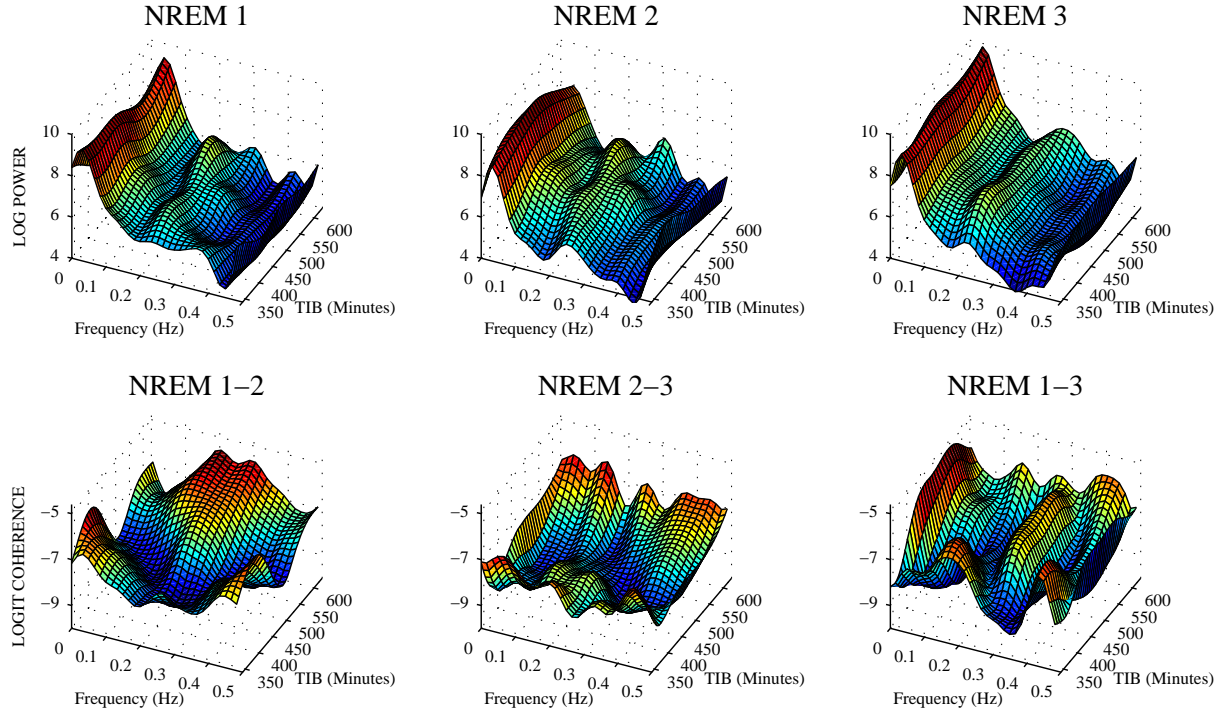


Figure 2: Estimated conditional log-spectra for each period of NREM (top panel) and estimated logit of conditional coherence between each period (bottom panel).

## 5.2 Analysis of Within-Period Power

We consider two collapsed measures of within-period power. In addition to HF previously defined in Equation (7), we also consider LF/HF as

$$f_p^{LF/HF}(u) = \left\{ \int_{.04}^{.15} f_{pp}(\omega, u) d\omega \right\} / \left\{ \int_{.15}^{.40} f_{pp}(\omega, u) d\omega \right\}.$$

Estimates and 95% pointwise credible intervals for these two measures as functions of TIB are displayed in Figure 3 for each period.

HF power is relatively constant across TIB during NREM 1, while participants with a TIB of less than 400 minutes have decreased HF power during NREM 2 and 3 compared to those who spend more time in bed. Further, those who have an exceedingly small TIB display increased LF/HF power during NREM sleep compared to those who spend more TIB, especially during NREM 3. These characteristics are indicative of heightened physiological arousal and psychological stress.



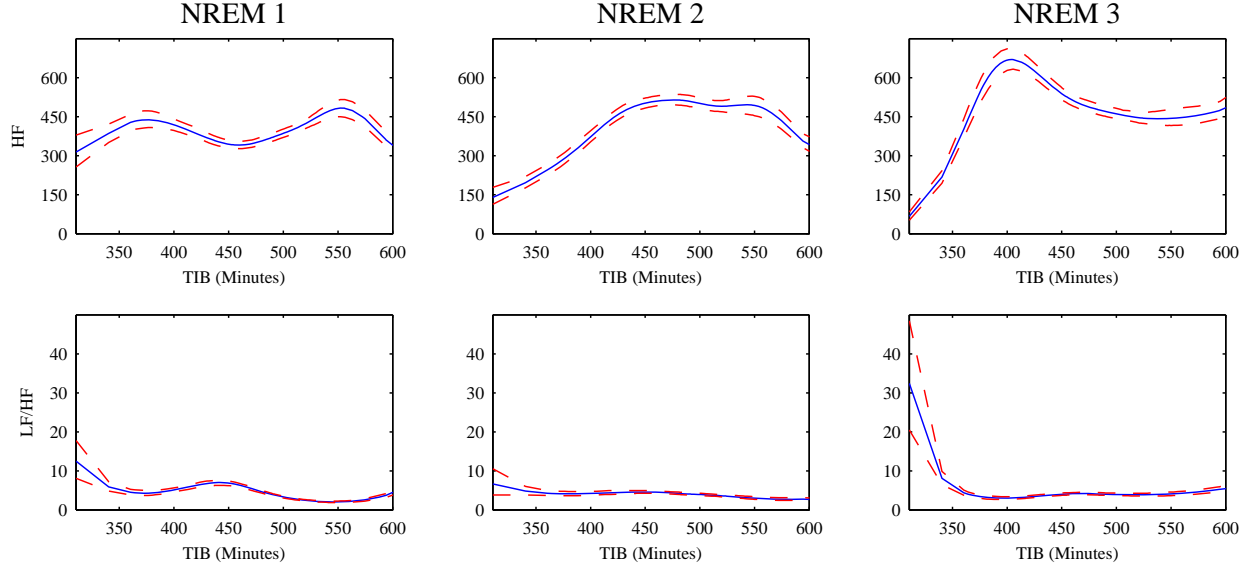


Figure 3: Estimated conditional HF (top panel),  $\hat{f}_p^{HF}$ , and LF/HF (bottom panel),  $\hat{f}_p^{LF/HF}$ , as functions of TIB with 95% pointwise credible intervals for each period of NREM.

Sleeping less than 7 hours per night has been shown to be associated with a multitude of negative health effects, including increased mortality (Buysse, 2014). The results of our analysis provide a potential pathway through which short sleep, which is inherently bounded by TIB, is connected to well-being: through increased stress and arousal towards the end of the night.

### 5.3 Analysis of Cross-Period Coherence

To investigate connections between cross-period coherence and TIB, we consider conditional HF band-squared coherence

$$\rho_{pq}^{2, HF}(u) = \left| \int_{.15}^{.40} f_{pq}(\omega, u) d\omega \right|^2 / \{f_p^{HF}(u) f_q^{HF}(u)\}$$

and display estimates on the logit scale,  $\text{logit}(\rho_{pq}^{2, HF}) = \log[\rho_{pq}^{2, HF} / (1 - \rho_{pq}^{2, HF})]$ , in the top panel of Figure 4. To better understand how changes in TIB are associated with HF coherence, we also examine first derivatives,

$$D_{pq}^{HF}(u) = d[\rho_{pq}^{2, HF}(u)] / du,$$

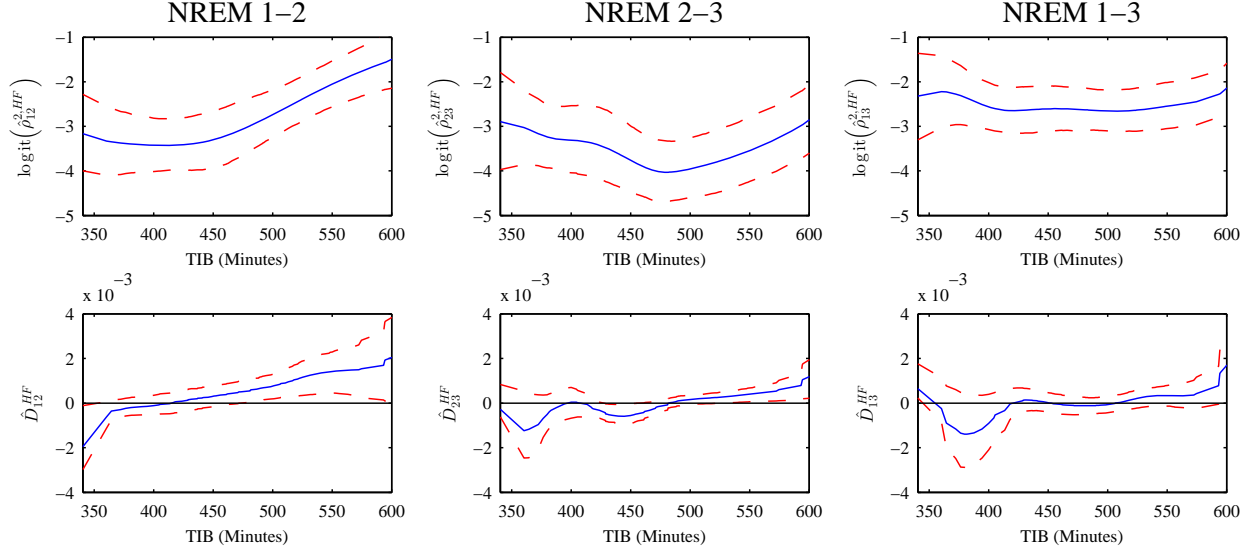


Figure 4: Estimated logit of integrated HF coherence (top panel),  $\text{logit}(\hat{\rho}_{pq}^{2, HF})$ , between each NREM period as functions of TIB and their first derivatives (bottom panel),  $\hat{D}_{pq}^{HF}$ , with pointwise 95% credible intervals.

whose estimates are displayed in the bottom panel of Figure 4. We find that the derivative of HF coherence between NREM 1 and 2 is positive for TIB greater than 500 minutes. This indicates that excessive increases in the amount of time spent in bed are associated with increased coherence in parasympathetic activity in the beginning of the night.

The relationship between excessive TIB and ill-health led Youngstedt and Kripke (2004) to propose modest sleep restrictions to increase quality of life and survival, especially for older adults, who tend to spend more time in bed as compared to younger adults. However, these restrictions must be used with great care as they can potentially lead to negative health effects (Reynolds III et al., 2010; Reynold et al., 2014). Our results demonstrate that excessive TIB is associated with a coherence in parasympathetic activity in the beginning of the night that is not present in moderate and short TIB. A possible explanation for this relationship is that extensive TIB can cause an increase in the amount of time spent awake while in bed, or lead to fragmented sleep. The roles of and relationships between physiological activity during different sleep cycles could change as sleep becomes more fragmented. These findings provide some of the first potential insights into the biological pathway through which excessive TIB can be connected to negative health, which can potentially be used to inform

optimal sleep restriction strategies in older adults.

## 6 Final Remarks

This article introduces a novel approach to analyzing associations between multiple time series and cross-sectional outcomes when data are observed from multiple subjects. A new measure of association, the conditional power spectrum, is introduced and its Cholesky components are modeled as bivariate functions of frequency and cross-sectional outcome. A MCMC algorithm is developed for model fitting allowing for inference on any function of the power spectrum. The method was motivated by a sleep study and uncovered connections between excessive time in bed and heightened arousal and stress that could not have been uncovered through traditional methods.

We conclude this section by discussing three extensions to the proposed methodology. First, the model is formulated to investigate the association between power spectra and a single cross-sectional variable. The model could easily be extended through higher-order tensor product models to include multiple variables, such as the amount of time it takes to fall asleep and the number of awakenings during the night. However, such a model would provide inference on the effect of these variables on the power spectrum conditional on the other variables, complicating interpretation when these variables are highly correlated. Future work will explore an interpretable canonical correlation type dimension reduction of a collection of correlated variables and multivariate spectral matrices, which can be viewed as a multivariate extension of Krafty and Hall (2013). Second, our application focused on HRV, due to the insights that it provides into autonomic nervous system activity. One could also explore the spectral analysis of other PSG channels, as well as the simultaneous coupling of channels. However, each channel of the PSG is sampled at a different rate. The second extension will develop conditional spectral analysis of time series with different sampling rates. Finally, since we were motivated by the analysis of HRV during epochs within NREM that are approximately stationary, we focused on stationary time series. For more highly sampled signals such as EEG, this assumption is not valid. A conditional time-frequency

analysis for signals that are locally stationary will also be explored.

## Supplementary Material

Supplementary material in the form of a pdf file and a zip file are available on the journal's website. The material in the pdf file contains two appendices: Appendix A provides results from simulation studies and Appendix B contains details concerning the sampling scheme. The a zip file contains a Matlab program for implementing the proposed procedure.

## References

- Burr, R. L. (2007), "Interpretation of normalized spectral heart rate variability indices in sleep research: A critical review," *Sleep*, 30, 913–919.
- Buysse, D. J. (2014), "Sleep health: Can we define it? Does it matter?" *Sleep*, 37, 9–17.
- Crainiceanu, C. M., Ruppert, D., and Wand, M. P. (2005), "Bayesian analysis of penalized spline regression using WinBUGS," *Journal of Statistical Software*, 14, 1–24.
- Dai, M. and Guo, W. (2004), "Multivariate spectral analysis using Cholesky decomposition," *Biometrika*, 91, 629–643.
- Eubank, R. L. (1999), *Nonparametric Regression and Spline Smoothing*, New York: Marcel Dekker, 2nd ed.
- Foley, D. J., Ancoli-Israel, S., Britz, P., and Walsh, J. (2004), "Sleep disturbances and chronic disease in older adults: Results of the 2003 National Sleep Foundation Sleep in America Survey," *Journal of Psychosomatic Research*, 56, 497–502.
- Foley, D. J., Monjan, A. A., Brown, S. L., Simonsick, E. M., Wallace, R. B., and Blazer, D. G. (1995), "Sleep complaints among elderly persons: An epidemiologic study of three communities." *Sleep*, 18, 425–432.

- Gelman, A. (2006), “Prior distributions for variance parameters (Comment on article by Browne and Draper),” *Bayesian Analysis*, 1, 515–534.
- Gu, C. (1992), “Cross-validating non-Gaussian data,” *Journal of Computational and Graphical Statistics*, 1, 169–179.
- (2013), *Smoothing Spline ANOVA Models*, New York: Springer-Verlag, 2nd ed.
- Gu, C. and Kim, Y.-J. (2002), “Penalized likelihood regression: general formulation and efficient approximation,” *Canadian Journal of Statistics*, 30, 619–628.
- Hall, M., Thayer, J., Germain, A., Moul, D., Vasko, R., Puhl, M., Miewald, J., and Buysse, D. (2007), “Psychological stress is associated with heightened physiological arousal during NREM sleep in primary insomnia,” *Behavioral Sleep Medicine*, 5, 178–193.
- Hall, M., Vasko, R., Buysse, D., Ombao, H., Chen, Q., Cashmere, J., Kupfer, D., and Thayer, J. (2004), “Acute stress affects heart rate variability during sleep,” *Psychosomatic Medicine*, 66, 56–62.
- Kimeldorf, G. S. and Wahba, G. (1970), “A correspondence between Bayesian estimation on stochastic processes and smoothing by splines,” *Annals of Mathematical Statistics*, 41, 495–502.
- Krafty, R. T. and Collinge, W. O. (2013), “Penalized multivariate Whittle likelihood for power spectrum estimation,” *Biometrika*, 100, 447–458.
- Krafty, R. T. and Hall, M. (2013), “Canonical correlation analysis between time series and static outcomes, with application to the spectral analysis of heart rate variability,” *Annals of Applied Statistics*, 7, 570–587.
- Malik, M., Bigger, J. T., Camm, A. J., Kleiger, R. E., Malliani, A., Moss, A. J., and Schwartz, P. J. (1996), “Heart rate variability - standards of measurement, physiological interpretation, and clinical use,” *Circulation*, 93, 1043–1065.

- Monk, T., Reynolds, C. F., Kupfer, D. J., Buysse, D. J., Coble, P. A., Hayes, A. J., MacHen, M. A., Petrie, S. R., and Ritenour, A. M. (1994), “The Pittsburgh Sleep Diary,” *Journal of Sleep Research*, 3, 111–120.
- Moulin, P. (1994), “Wavelet thresholding techniques for power spectrum estimation,” *IEEE Transactions on Signal Processing*, 42, 3126–3136.
- Nychka, D. and Cummins, D. (1996), “Comment on paper by Eilers and Marx,” *Statistical Science*, 11, 104–105.
- Pawitan, Y. and O’Sullivan, F. (1994), “Nonparametric spectral density estimation using penalized Whittle likelihood,” *Journal of the American Statistical Association*, 89, 600–610.
- Qin, L. and Wang, Y. (2008), “Nonparametric spectral analysis with applications to seizure characterization using EEG time series,” *Annals of Applied Statistics*, 2, 1432–1451.
- Reynold, A. M., Bowles, E. R., Saxena, A., Fayad, R., and Youngstedt, S. D. (2014), “Negative effects of time in bed extensions: A pilot study,” *Journal of Sleep Medicine and Disorders*, 1, 1002.
- Reynolds III, C. F., Serody, L., Okun, M. L., Hall, M. H., Houck, P. R., Patrick, S., Maurer, J., Bensasi, S., Mazumdar, S., Bell, B., Nebes, R. D., Miller, M. D., Dew, M. A., and Nofzinger, E. A. (2010), “Protecting sleep, promoting health in later life: A randomized clinical trial,” *Psychosomatic Medicine*, 72, 178–186.
- Rosen, O. and Stoffer, D. (2007), “Automatic estimation of multivariate spectra via smoothing splines,” *Biometrika*, 94, 335–345.
- Rosen, O., Stoffer, D. S., and Wood, S. (2009), “Local Spectral Analysis via a Bayesian Mixture of Smoothing Splines,” *Journal of the American Statistical Association*, 104, 249–262.
- Rosen, O., Wood, S., and Stoffer, D. (2012), “AdaptSPEC: Adaptive Spectral Estimation

- for Nonstationary Time Series,” *Journal of the American Statistical Association*, 107, 1575–1589.
- Shumway, R. H. and Stoffer, D. S. (2011), *Time Series Analysis and Its Applications: With R Examples*, New York: Springer, 3rd ed.
- Siegel, J. M. (2005), “Clues to the functions of mammalian sleep,” *Nature*, 437, 1264–1271.
- Speckman, P. L. and Sun, D. (2003), “Fully Bayesian spline smoothing and intrinsic autoregressive priors,” *Biometrika*, 90, 289–302.
- Stoffer, D. S., Han, S., Qin, L., and Guo, W. (2010), “Smoothing spline ANOPOW,” *Journal of Statistical Planning and Inference*, 140, 3789–3796.
- Wahba, G. (1978), “Improper priors, spline smoothing and the problem of guarding against model errors in regression,” *Journal of the Royal Statistical Society, Series B*, 40, 364–372.
- Wand, M., Ormerod, J., Padoan, S., and Frühworth, R. (2012), “Mean field variational Bayes for elaborate distributions,” *Bayesian Analysis*, 7, 847–900.
- Wang, J.-L., Chiou, J.-M., and Müller, H.-G. (2016), “Review of Functional Data Analysis,” *Annual Review of Statistics and Its Application*, 3, In Press.
- Whittle, P. (1953), “Estimation and information in stationary time series,” *Arkiv for Matematik*, 2, 423–434.
- (1954), “Some recent contributions to the theory of stationary processes,” *A Study in the Analysis of Stationary Time Series*, 2, 196–228.
- Wood, S. A., Jiang, W., and Tanner, M. (2002), “Bayesian mixture of splines for spatially adaptive nonparametric regression,” *Biometrika*, 89, 513–528.
- Wood, S. N. (2006), *Generalized Additive Models: An Introduction with R*, Boca Raton: CRC Press.
- Youngstedt, S. D. and Kripke, D. F. (2004), “Long sleep and mortality: Rationale for sleep restriction,” *Sleep Medicine Review*, 8, 159–174.